DOCUMENT RESUME

ED 061 283 TM 001 234

AUTHOR Schnittjer, Carl J.
TITLE The Use of Linear Programming for Prediction.
PUB DATE Apr 72
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, April 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Comparative Analysis; Educational Administration; Graduate Students; *Linear Programing; Mathematical Applications; *Mathematical Models; *Multiple Regression Analysis; *Prediction; Predictive Ability (Testing); Predictive Measurement; Predictive Validity; Predictor Variables; Scores

ABSTRACT

The purpose of the study was to develop a linear programming model to be used for prediction, test the accuracy of the predictions, and compare the accuracy with that produced by curvilinear multiple regression analysis. (Author)

ED 061283

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE USE OF LINEAR PROGRAMMING FOR PREDICTION

By

Carl J. Schnittjer

A paper presented at the
1972 Annual Meeting of the
American Educational Research Association

April 1972

TM 001 234

INTRODUCTION.

The purpose of the study was to develop a linear programming model to be used for prediction, test the accuracy of the predictions, and compare the accuracy with that produced by curvilinear multiple regression analysis.

THE LINEAR PROGRAMMING MODEL.

The discussion presumes that the reader has some background in linear programming. Detailed discussions of linear programming theory may be found in books by Hadley, Hillier, Wagner, and Van Dusseldorp.

The linear programming model developed in this study was used to assign weights to the various amounts or levels within each variable. The objective was to assign these weights so that the resulting predicted scores produced would be as close as possible to the actual scores being predicted for each person. Thus the objective function was the sum of the residuals:

$$R(1) + R(2) + R(3) + \ldots + R(n), \quad R(i) = |A(i) - P(i)|$$

where, R(i) = absolute value of the residual for person i.

A(i) = actual criterion score for person i.

P(i) = predicted score for person i.

The "objective" was to minimize the objective function subject to three kinds of constraints:

1. Person Equations: These linear equations essentially related each person to the amount of each variable he possessed.

2. Variable Contribution Limiting Equations: These linear equations put upper and lower bounds on the amount that any variable might contribute to the predicted score.

3. Level Width Restriction Equations: These equations were used to restrict the consecutive levels within each variable. They could be used to force each variable to be monotonic increasing or decreasing so that each successive level would contribute more(or less) then its preceding level.

TESTING THE LINEAR PROGRAMMING MODEL.

The linear programming model for prediction was tested using data on 66 Ph.D. graduates in educational administration at the University of Iowa. The following independent variables were used:

1. Age at admission to the Ph.D. program(AGE).
2. Undergraduate grade point average(GPAU).
3. Graduate grade point average at time of admission to the Ph.D. program(GPAG).
4. Verbal score on the Graduate Record Exam(GREV).
5. Quantitative score on the Graduate Record Exam(GREQ).
6. Number of years elapsed, B.A. degree to M.A. degree(YRSM).
7. Number of years elapsed, M.A. degree to Ph.D. degree(YRSP).
8. Number of years of teaching experiance(YRST).
9. Number of years of administrative experiance(YRSA).
10. Number of graduate hours earned prior to admission to the Ph.D. program(HRS).

The following dependent(criterion) variables were used in the study:

1. Final graduate grade point average(GRADGPA).
2. Professors' general rating of the student(GNRLRTNG).

3. Professors' rating of the student on the comprehensive exam(COMPRTNG).

4. Professors' rating of the students thesis(THESRTNG).

The independent variables were divided into intervals or levels, with the number of intervals for each variable ranging from five(5) to 27.

The sums of residuals produced by the linear programming model were compared to those produced by the usual curvilinear multiple regression analysis. Also, a simulation-prediction procedure was developed in which 53 sets of 50 persons each were randomly chosen from the 66 persons in the population. The linear programming model was then run with each group of 50 randomly selected persons and the results tested for prediction potential on the 16 persons not included in that particular run.

RESULTS PRODUCED BY THE MODEL.

Once the model was constructed, a number of adjustments were made to enhance its performance while at the same time preserving the basic structure.

The results produced by the model for each of the four dependent variables are shown in Table 1. The sums of residuals produced may then be compared to those for curvilinear multiple regression analysis as shown in Table 2.

TABLE 1

Sum of Residuals from Linear Programming Regression Analysis

Criterion Variable	Lower Bound	Upper Bound	Level Width	Monotonic Variables: Increase	Monotonic Variables: Decrease	Sum of Residuals
GRADGRP	-100	100	fixed	all		972.04
GNRNRTNG	-100	100	fixed	all		2865.18
COMPRTNG	-100	100	fixed	all		3618.15
THESRTNG	-100	100	fixed	all		3142.02
GRADGPA	-100	100	≥ 0	all		661.47
GNRLRTNG	-100	100	≥ 0	all		2222.00
COMPRTNG	-100	100	≥ 0	all		2817.78
THESRTNG	-100	100	≥ 0	all		2556.00
GRADGPA	-100	100	≥ 0	2-6,8	1,7,9,10	640.08
GNRLRTNG	-100	100	≥ 0	2-6,8	1,7,9,10	11716.25
COMPRTNG	-100	100	≥ 0	2-6,8	1,7,9,10	2269.74
THESRTNG	-100	100	≥ 0	2-6,8	1,7,9,10	2241.50
GRADGPA	-100	100	≥ 0	2-6,8		343.78
GNRLRTNG	-100	100	≥ 0	2-6,8		906.96
COMPRTNG	-100	100	≥ 0*	2-6,8		1035.89
THESRTNG	-100	100	≥ 0*	2-6,8		1121.03
GRADGPA	-inf.	+inf.	free			0
GNRLRTNG	-inf.	+inf.	free			0
COMPRTNG	-inf.	+inf.	free			0
THESRTNG	-inf.	+inf.	free			0

*Variables 1,7,9, and 10 are free (i.e., have no limitations imposed on level width or contribution).

TABLE 2

Transformations Used in the "least Squares" Curvilinear Multiple Regression Analysis with the Resulting Betas, R^2, and Sums of Residuals .

#	*	VARIABLE										R^2	Sum of Residuals
		AGE	GPAU	GPAG	GREV	GREQ	YRSM	YRSP	YRST	YRSA	HRS		
1	T	Ln	Square	Cube	Ln	Cube	Cube	none	--	Ln	Ln	.497	950.0
	B	-8.463	.0341	.00086	12.294	.00005	.00005	3.311	--	-289.9	-75.65		
2	T	Ln	Square	Cube	Square	Cube	Sq. Rt.	Cube	Ln	--	Ln	.438	2694.2
	B	-.4211	.30726	.02082	.12082	.10519	.22589	.23605	.15679	--	-.2423		
3	T	Sq. Rt.	Cube	Cube	none	Cube	Cube	Cube	Cube	Cube	Cube	.318	3543.4
	B	-20.70	.00034	.00122	1.394	.00001	.00029	-.00018	.00070	.00027	-.00012		
4	T	none	Ln	Cube	Ln	Cube	Cube	Cube	Sq. Rt.	Ln	Cube	.365	2924.9
	B	-17.09	-15.49	.00142	133.27	.00023	.00074	.00097	247.71	-1108.	-.00008		

Criterion variables: 1 = GRADGPA, 2 = GNRLRTNG, 3 = COMPRTNG, 4 = THESRTNG.
*T = Transformation, B = Beta Weight.

DISCUSSION OF THE RESULTS.

It would appear from examination of the data that linear programming does offer a viable alternative to the usual "least squares" regression analysis for purposes of prediction. This conclusion can be supported by the following evidence:

1. It was clearly demonstrated that linear programming can be applied to prediction problems.
2. The data presented in Tables 1 and 2 show that the linear programming model produced smaller sums of residuals than the "least squares" regression analysis in all but the most restrictive variation of the model.
3. The least restrictive variation of the linear programming model, that composed of only "person equations", produced a sum of residuals of zero. This finding is felt to be very significant.
4. The model can be adjusted to conform to as weak or as tight a set of restrictions as desired. Such flexibility is clearly an advantage.

While this study was concerned with prediction of success, it seems reasonable to conclude that the results of this study might well apply to regression problems in general. The ability to produce a zero sum of residuals in itself would indicate the possibility of real potential in this regard.

While the potential of this application has been demonstrated in the case of this prediction problem, the practical considerations for application of the linear programming model need further exploration. As with any such mathematical procedure, the application of the proposed linear programming model to decision making must be weighted in the context of the situation, the accuracy required, and the accompanying consequences of the decision.

REFERENCES.

Bruno, James E. "A Linear Programming Approach to Position-Salary Evaluation in School Personnel Administration," Santa Monica: RAND Corporation, 1968, (Paper No. P-4039).

Draper, N. W., and Smith. Applied Regression Analysis. New York: John Wiley & Sons, 1966.

Hadley, George. Linear Programming. New York: Addison-Wesley, 1962.

Hillier, Frederick S. and Gerald J. Lieberman. Introduction to Operations Research. San Francisco: Holden-Day, Inc., 1967.

Karst, Otto J. "Linear Curve Fitting Using Least Deviations," American Statistical Association Journal, 53:118. March, 1958.

Rehmus, Frederick P., and Harvey M. Wagner. "Applying Linear Programming to Your Pay Structure," Business Horizons, 4:89, Winter, 1963.

Schnittjer, Carl J. "Prediction of Success Through Linear Programming," Unpublished Doctoral dissertation, University of Iowa, Iowa City, Iowa, 1971.

Van Dusseldorp, Ralph A., Duane E. Richardson, and Walter J. Foley. Educational Decision-Making Through Operations Research. Boston: Allyn and Bacon, Inc., 1971.

Wagner, Harvey M. "Linear Programming Techniques for Regression Analysis," Journal of the American Statistical Association, 57:206-12, March, 1959.

Wagner, Harvey M. Principles of Operations Research. Englewood Cliffs: Prentice-Hall Inc., 1969.